Sparse-view Cone Beam CT Reconstruction using Data-consistent Supervised and Adversarial Learning from Scarce Training Data

Gabriel Maliakal†, Anish Lahiri*, Marc L. Klasky†, Jeffrey A. Fessler* *Fellow, IEEE*, and Saiprasad Ravishankar‡ *Senior Member, IEEE*

***Abstract*—Reconstruction of CT images from a limited set of projections through an object is important in several applications ranging from medical imaging to industrial settings. As the number of available projections decreases, traditional reconstruction techniques such as the FDK algorithm and model-based iterative reconstruction methods perform poorly. Recently, data-driven methods such as deep learning-based reconstruction have garnered a lot of attention in applications because they yield better performance when enough training data is available. However, even these methods have their limitations when there is a scarcity of available training data. This work focuses on image reconstruction in such settings, i.e., when both the number of available CT projections and the training data is extremely limited. We adopt a sequential reconstruction approach over several stages using an adversarially trained shallow network for 'destreaking' followed by a data-consistency update in each stage. To deal with the challenge of limited data, we use image subvolumes to train our method, and patch aggregation during testing. To deal with the computational challenge of learning on 3D datasets for 3D reconstruction, we use a hybrid 3D-to-2D mapping network for the 'destreaking' part. Comparisons to other methods over several test examples indicate that the proposed method has much potential, when both the number of projections and available training data are highly limited.**

***Index Terms*—Sparse-views, Computed tomography, Machine learning, Deep learning, Image reconstruction.**

I. Introduction

Computed Tomography (CT) is an important imaging modality across applications in medicine, industry, science and security. In this work, we develop an iterative machine learning-based approach for 3D cone beam CT reconstruction from very limited measurements or projections, and using limited training data. In several applications, acquiring many projections or 'views' through the object may be undesirable or impossible. This constraint may be to reduce exposure to radiation in medical imaging applications, or due to only preset limited or sparse views being possible in industrial or security applications. Moreover, in dynamic imaging applications, where the object is changing while being imaged, we would also be limited to fewer views per temporal state, to prevent blurring. Deep learning-based approaches have the potential to perform better in this scenario, but still perform poorly when there is a scarcity of available data for training, such as in national security applications where experimental data is limited and accurate simulations are expensive [1]. While there are approaches that reconstruct from very limited projections, they either do not target 3D CBCT imaging [2]–[5], or rely upon many paired training image volumes [6]–[11].

*A. Lahiri and J. A. Fessler are with the Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI 48014 USA. Emails: anishl@umich.edu, fessler@umich.edu.

‡S. Ravishankar is with the Department of Computational Mathematics, Science and Engineering, and the Department of Biomedical Engineering, Michigan State University, East Lansing, MI 48824 USA. Email: ravisha3@msu.edu.

†M. L. Klaksy is with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA. Email: mklasky@lanl.gov.

†Gabriel Maliakal is with the Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824 USA, and at the time of writing the paper was an intern at Los Alamos National Laboratory, Los Alamos, NM 87545 USA. Email: maliakal@msu.edu.

II. Algorithm & Problem Setup

Our proposed method for CBCT reconstruction focuses on addressing two primary challenges: a very limited number of available views, and limited number of available training objects. We address the former through a combination of three aspects: (1) using an edge-preserving regularized reconstruction [12] to initialize our iterative-type algorithm; (2) including an adversarial component to the training loss function for our learned destreaking networks (similar to generative adversarial networks or GANs); and (3) including data-consistency blocks that reinforce acquired measurements in the destreaked 3D reconstruction.

The problem of scarce training data is addressed primarily by two approaches. First, we split an entire image volume into patches in the form of overlapping subvolumes. Essentially, this step localizes the scope of CNN-based destreaking to a comparatively smaller neighborhood, while allowing us to generate many training examples from a single image volume. Second, we use a shallow destreaking CNN to avoid overfitting to the training data. To reduce computation time associated with multiple 3D convolutions and subsequent patch aggregation, the CNN is designed to map 3D subvolumes to 2D slices [13]. This approach enables using 3D contextual information for the destreaking task, while removing the need for patch averaging (of overlapping 3D patches) and associated artifacts during aggregation.

Many learning-based reconstruction approaches in applications like MRI work through end-to-end training, whereas such approaches are less practicle in CT due to the complexity of the system matrix. Thus, our developed algorithm operates

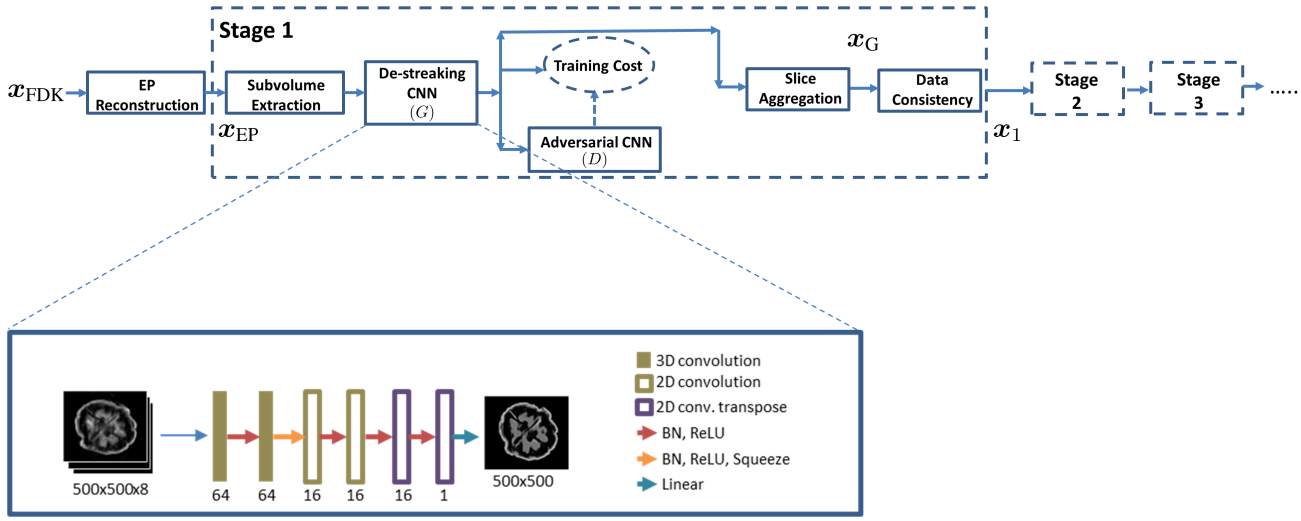


Figure 1. Flow diagram depicting the overall pipeline of our algorithm, where $\boldsymbol{x}_{\text{FDK}}$ is the FDK reconstruction, $\boldsymbol{x}_{\text{EP}}$ is an edge-preserving regularized reconstruction, $\boldsymbol{x}_{\text{G}}$ is the generator's output after slice aggregation, and $\boldsymbol{x}_1$ is the output of the first stage.

as a multi-stage greedy approach similar to works like [14], [15]. Each stage is composed of a CNN that maps each 3D subvolume with streaking artifacts to a clean 2D slice corresponding to the slice at the centre of the subvolume. Because the objects considered here have finite support, we treat the slices at the edge of the volume in the direction of aggregation are being empty, and set them to zero. One could use other boundary conditions for long objects [16]. Once an entire image volume has been aggregated from individual clean slices, this volume is passed through a data-consistency update to reinforce acquired measurements and reduce any 'hallucinations' introduced by the network. This output subvolume is then provided as input to the next stage. Fig. 1 depicts the process that is akin to algorithm unrolling [17]. As mentioned earlier, to reduce noise and streaks, the input to the first stage of our method is an edge-preserving regularized (iteratively obtained) reconstruction, using the regularizer in [12] and the algorithm in [18], which in turn is initialized with an FDK reconstruction for faster convergence.

We train the CNN parameters separately for each stage. The training loss for the destreaking CNN in each stage consists of a weighted combination of a masked mean squared error term calculated over a region of interest using the ground truth training image slices, and an adversarial input from another CNN that acts as a discriminator for the output of the destreaking CNN (also specific to the stage). Adversarial training often is posed as a min-max optimization problem [19], but the practical implementation involves alternating between updating generator (destreaking) network G parameters ϕ^k and discriminator network D parameters θ^k, with more frequent updates of the generator parameters. Our approach to updating the generator weights for the kth stage is mathematically expressed as:

$$\begin{aligned}\hat{\theta}^k = \arg\min_{\theta} \; & -\lambda \mathbb{E}\big[D_{\phi^k}(G_\theta(\mathcal{P}_3 \boldsymbol{x}_{k-1}))\big] \\ & + \mathbb{E}\big[\|\mathcal{P}_3^{2,\text{mid}} \boldsymbol{x}_{\text{GT}} - G_\theta(\mathcal{P}_3 \boldsymbol{x}_{k-1})\|_2^2\big],\end{aligned} \tag{1}$$

where $\boldsymbol{x}_{k-1}$ is the output of the $(k-1)$th stage of our algorithm, $\boldsymbol{x}_0$ is set to be $\boldsymbol{x}_{\text{EP}}$ (the edge-preserving regularized reconstruction), $\boldsymbol{x}_{\text{GT}}$ is the ground truth, λ is a regularization parameter that varies as the weights of G^k are updated, $\mathcal{P}_3$ is a 3D patch or subvolume extraction operator, and $\mathcal{P}_3^{2,\text{mid}}$ is an operator that extracts the 2D central slice from an image subvolume, where the position of the subvolume is determined by $\mathcal{P}_3$. We restricted the reconstructions to a region in the image volume containing the object of interest. The expectation $\mathbb{E}$ is taken over the set of training examples.

Our approach to updating the the discriminator network parameters is likewise given as:

$$\begin{aligned}\hat{\phi}^k = \arg\min_{\phi} \; & \mathbb{E}\big[\big(D_\phi(G_{\theta^k}(\mathcal{P}_3 \boldsymbol{x}_{k-1})) - 0\big)^2\big] \\ & + \mathbb{E}\big[\big(D_\phi(\mathcal{P}_3^{2,\text{mid}} \boldsymbol{x}_{\text{GT}}) - 1\big)^2\big].\end{aligned} \tag{2}$$

The data-consistency update involves seeking an image that is consistent with the acquired measurements while still being 'close' to the slice-aggregated destreaked image. The optimization problem for this step is framed as:

$$\boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{x} - \boldsymbol{x}_{k,\text{G}}\|_2^2, \tag{3}$$

where $\boldsymbol{A}$ is the CBCT system matrix, implemented with the separable-footprint projector [20], $\boldsymbol{y}$ denotes the projections or acquired measurements, $\beta > 0$ is a regularization parameter, and $\boldsymbol{x}_{k,\text{G}}$ is the output of the generator after slice aggregation at the kth stage. We used an ordinary least-squares (LS) data-fit term rather than a weighted LS (WLS) term because the focus here is on sparse views rather than low-dose imaging, but the method generalizes directly to the WLS case. We used 50 conjugate gradient (CG) iterations to minimize (3).

III. Methods

A. Dataset and Experimental Settings

To train and test our method, we examined two diverse problems. The first problem is a standard sparse-view image reconstruction problem in which we utilized the publicly available 3D walnut CT dataset [21]. To study the ability

to learn from very limited data, we used a single walnut for training our method, and tested our algorithm on 5 different walnuts. Furthermore, extremely limited data with 8 or 4 views/projections through the walnuts were used in training and testing the network. Separate networks were trained for reconstructing image volumes from 4 and 8 views, respectively. These CBCT views were generated using the MIRT [22] package, and were equally spaced over 360 degrees. The distance from source to detector was set to be 20 cm, the distance from the object to the detector was 4.08 cm, and each projection view was 150×150 pixels of size ≈ 0.4 mm^2

Our second problem examines [23] the reconstruction of complex hydrodynamic phenomena using the computational fluid dynamics software CTH [24]. Specifically, we examined the ability to perform reconstructions from a single hydrodynamic training simulation using only a single temporal slice and 8 projections. To illustrate the variation of the voxel values, Fig. 4 provides a few one dimensional profiles of different simulations across the central row of the central slice.

The ability of the algorithm to reconstruct hydrodynamic simulations generated with different hydrodynamic parameters was tested using two simulations, one referred to as simulation 12 (see Fig. 5 and Fig. 6) that was generated using one set of parameters at an early time and another referred to as simulation 7 that was generated using a different set of parameters and at a later time in its simulation. We also used two grossly different volumes to train the network to examine the ability of the network to generalize from different training cases. The CBCT data here were generated in the same manner as the walnut data. The 2.5D Deep CNN and our proposed method were both trained on the volume within a bounding box containing the outermost shell; this outer shell can be seen in Fig. 5. For comparison, all reconstruction metrics were reported in a $130 \times 130 \times 130$ bounding box containing the inner topological structures in the volume. This box is shown in the zoomed in parts of images shown in Fig. 5.

Because the CBCT system simulated here has a small cone angle (almost parallel beam), 8 views over 360° probably has only a bit more information than 4 views over 180°. The image volume for each walnut was $501 \times 501 \times 501$. The dimensions of each voxel were approx. $0.12 \times 0.12 \times 0.12$ mm^3, whereas the dimensions of the hydrodynamic simulations were $448 \times 448 \times 448$ with voxel size $0.025 \times 0.025 \times 0.025$ cm^3.

B. Compared Methods

To assess the performance of our method, we used 4 stages of our proposed algorithm to achieve a compromise between image quality and runtime. We compared the output for all 5 test walnuts to the conventional FDK reconstruction, an edge-preserving (EP) regularized (MBIR) reconstruction [12], as well as the slice-aggregated output from a single stage of our destreaking CNN without data consistency. We also compared our proposed method to a deep 2.5D (residual) CNN whose architecture is as in [13] and without data consistency. This network had 16 hidden convolutional layers, whose input was a subvolume of three slices, and its output was a single slice trained to be the residual needed to be added to the middle slice of the EP reconstruction input to get the full-views image. The network was trained with a learning rate of 0.001 for Walnuts and 0.0002 for hydrodynamic simulations, both using the Adam optimizer with a learning rate scheduler that multiplied the current learning rate by 0.1 every 50 epochs. The network was trained for 100 epochs. A batch size of 4 was used for Walnut dataset whereas a batch size of 6 was used for hydrodynamic simulations. This network was trained using masked mean squared error between output residual and ground truth-based residual, where the mask was the bounding box containing the edges of the training volume. Ground truth residuals were obtained by subtracting the EP slice from full-views or ground truth image. We trained this model with and without data augmentation that involved random elastic deformations, random rotations, and flips. However, we observed that data augmentation did not improve the performance on the unseen test volumes for both the datasets. Data augmentation increased the training time, but it still gave similar performance to without data augmentation. We therefore simply show the latter results. Finally, we also compare the robustness of our proposed four stage method by testing with smaller object to detector and source to detector distances, wherein cone beam artifacts are more prevalent and different than in the training set. [1]

C. Performance Metrics

We primarily used the normalized mean absolute error (NMAE) as a metric for evaluating the performance of various methods. For the walnut data, the error is evaluated over the voxels within the region-of-interest (ROI) of a three-dimensional mask obtained by dilating a ground truth segmentation of the walnut being reconstructed. The masked region includes all voxels within the shell of the walnut. The NMAE normalization used the mean intensity of the ground truth voxels within this mask. Essentially, $E_{\text{NMAE}}(x_{GT}, x_O, \mathcal{M}) = \|\mathcal{M} \odot (x_{GT} - x_O)\|_1 / \|\mathcal{M} \odot x_{GT}\|_1$, where x_{GT} is the ground truth image volume, x_O is the reconstruction whose quality is being evaluated, and $\mathcal{M}$ is a binary mask specific to the test volume (e.g., which excludes any pixel not within the dilation of the outer boundary for walnuts[2]). For computing metrics for hydrodynamic data, we let the mask cover a central region of the volume having the predominant object dynamics. The masks were applied so that a cubical region of size $130 \times 130 \times 130$ encompasses the central toplogical features. This was done to evaluate the reconstructions of this interior region which had complex features of interest.

Another metric that is used for comparison in our work is the normalized high-frequency error norm or NHFEN [25]. We computed the HFEN for every slice of the reconstructed walnut as the ℓ_2 norm of the difference of masked edges (obtained through a high-pass filtering) between the input and reference images. The masking is done similarly as described earlier. A Laplacian of Gaussian (LoG) filter was used as the edge

[1]Implementation of proposed and compared methods are available in https://github.com/gtm2122/SparseViewCT-TCI

[2]These are obtained by a histogram-based thresholding of the corresponding ground truth volumes for the test walnuts.

detector. The kernel size was set to 15×15, with a standard deviation of 1.5 pixels. The normalization was performed over the high frequency components of the ground truth image over the masked ROI. Mathematically, this metric is calculated as $E_{\text{NHFEN}}(x_{GT}, x_O, \mathcal{M}) = \frac{1}{N}\sum_i \|\mathcal{H}(\mathcal{M}[:,:,i] \odot x_{GT}[:,:,i]) - (\mathcal{H}(\mathcal{M}[:,:,i] \odot x_O[:,:,i])\|_2 / \|\mathcal{H}(\mathcal{M}[:,:,i] \odot x_{GT}[:,:,i])\|_2$, where $\mathcal{H}$ denotes the LoG filter described earlier, i indexes the slices of the image volume in the z direction, where N is the total number of slices in that direction, and the other symbols have their usual meaning, as described previously. An advantage of using such normalized metrics is that it allows for the evaluation of the reconstruction quality only in areas of interest in the volume, disregarding the effect of empty spaces around it.

IV. Results

A. Walnut Dataset Reconstruction Comparisons

Table I and Table II compare the reconstruction performance of various methods (including our own) described in the previous sections. The proposed approach substantially improves the NMAE and NHFEN compared to the reference methods for 8 and 4 acquired projections, respectively. As expected, the quality of reconstructions using 4 acquired projections was worse than when 8 projections were acquired for reconstruction.

Additionally, to better test the advantage of having multiple stages in our method, we trained the destreaking network in Stage 1 for 4 times the epochs as our multi-stage scheme, and followed up the network with data-consistency once. This method was compared against our proposed approach where each destreaking CNN was trained for 40 epochs followed by a data consistency step for a total of 4 stages. Table I shows this comparison in terms of reconstruction metrics. Fig. 2 shows a visual comparison of reconstruction quality for 8 views with the previously mentioned methods and our 4 stage approach. The proposed algorithm provides significantly higher quality reconstructions than the other methods. This is particularly evident in the extent to which our algorithm is able to restore the finer features of the walnuts, and has fewer artifacts.

Fig. 3 compares the reconstructions from 4 views. The quality of the reconstruction is poorer compared to that using 8 views, though the proposed approach still visibly outperforms the other methods.

B. Hydrodynamic Simulation Reconstruction Comparisons

Table III displays a similar comparison with two versions of our proposed model, each trained on a different hydrodynamic simulation using the same acquisition geometry as used for the walnuts in Table I. There were 12 simulations in total generated using 4 set of simulation parameters and each were sampled at different timepoints. Out of these 12, simulations 7 and 12 were used for training. These were generated using different sets of parameters and the training volumes corresponded to two different time points in the simulations. The results show that the proposed method outperforms the compared methods across a variety of test cases in both training scenarios. Fig. 5 compares reconstructions of a test simulation by various methods mentioned above; our proposed method clearly outperformed all compared methods. Fig. 6 compares reconstructions by our proposed method against 2.5D DnCNN [13] for a case where the latter showed a comparable reconstruction to our proposed method in terms of NMAE. Despite the similar NMAE, the image features are still visually better for our proposed network.

V. Conclusions and Future Work

This paper developed a method to provide high quality reconstructions from extremely limited CBCT projections and scarce training data. The key features of our approach were the multi-stage approach of alternating between learning-based destreaking and data consistency and the use of subvolume-based learning and shallower (adversarially trained) CNNs to combat over-fitting. In the future, we will focus on extending our method to dynamic imaging applications in CT, as well as being able to jointly segment and reconstruct three-dimensional objects. Towards this end, we are also interested in finding a better metric than the mean absolute error to assess the fidelity and quality of our reconstructions for specific tasks.

References

[1] M. Hossain, B. T. Nadiga, O. Korobkin, M. L. Klasky, J. L. Schei, J. W. Burby, M. T. McCann, T. Wilcox, S. De, and C. A. Bouman, "High-precision inversion of dynamic radiography using hydrodynamic features," 2021. [Online]. Available: http://arxiv.org/abs/2112.01627

[2] Y. Han, J. Kang, and J. C. Ye, "Deep learning reconstruction for 9-view dual energy CT baggage scanner," in *Proc. 5th Intl. Mtg. on Image Formation in X-ray CT*, 2018, pp. 407–10.

[3] G. T. Herman and R. Davidi, "Image reconstruction from a small number of projections," *Inverse Prob.*, vol. 24, no. 4, p. 045011, Aug. 2008.

[4] E. T. Quinto, "Exterior and limited-angle tomography in non-destructive evaluation," *Inverse Problems*, vol. 14, no. 2, p. 339, 1998.

[5] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[6] M. Xiang, "Deep learning-based reconstruction of volumetric CT images of vertebrae from a single view X-ray image," 2020, theses and Dissertations. 2629. University of Wisconsin Milwaukee. [Online]. Available: https://dc.uwm.edu/etd/2629

[7] R. Anirudh, H. Kim, J. J. Thiagarajan, K. A. Mohan, K. Champley, and T. Bremer, "Lose the views: limited angle CT reconstruction via implicit sinogram completion," in *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*, 2017, pp. 6343–52.

[8] H. Kim, R. Anirudh, K. A. Mohan, and K. Champley, "Extreme few-view CT reconstruction using deep inference," *arXiv preprint arXiv:1910.05375*, 2019.

[9] E. Eulig, J. Maier, M. Knaup, N. R. Bennett, K. Hörndler, A. S. Wang, and M. Kachelrieß, "Deep learning-based reconstruction of interventional tools and devices from four X-ray projections for tomographic interventional guidance," *Medical physics*, vol. 48, no. 10, pp. 5837–5850, 2021.

[10] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 10 619–10 628.

[11] J. C. Montoya, C. Zhang, Y. Li, K. Li, and G.-H. Chen, "Reconstruction of three-dimensional tomographic patient models for radiation dose modulation in CT from two scout views using deep learning," *Medical Physics*, vol. 49, no. 2, pp. 901–916, 2022.

[12] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 678–89, Feb. 2015.

[13] A. Ziabari, D. H. Ye, S. Srivastava, K. D. Sauer, J. Thibault, and C. A. Bouman, "2.5D deep learning for CT image reconstruction using A multi-GPU implementation," in *asccs*, 2018, pp. 2044–9.

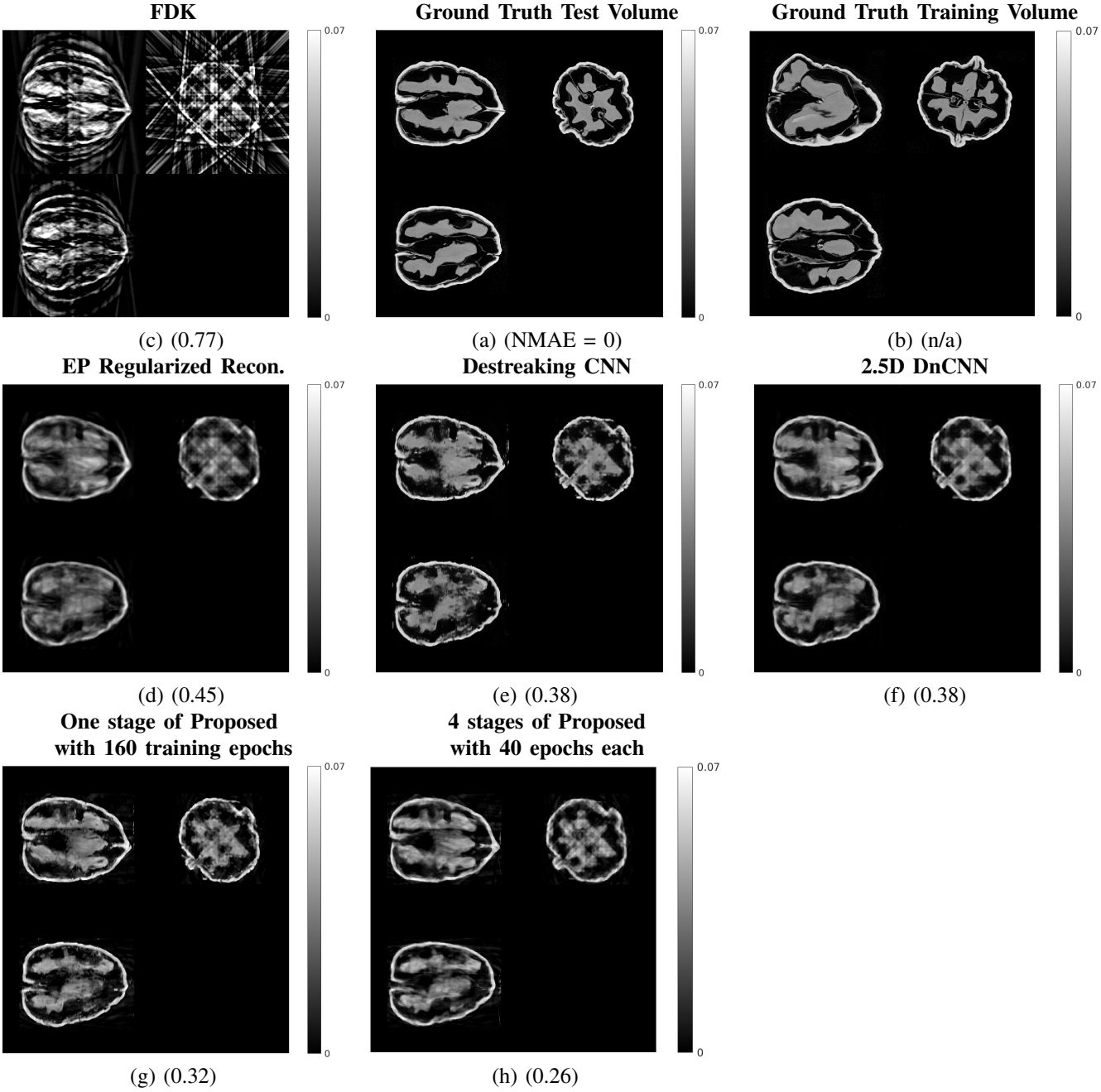


Figure 2. Comparison of the quality of reconstruction of our proposed algorithm (h) for walnut 2 (8 views) in Table I to various reference methods. Each subfigure depicts slices through the center of the walnut volume in three different directions (or sagittal, coronal and transverse orientations). The normalized mean absolute errors have also been shown underneath each subfigure. The central slices corresponding to the ground truth training walnut volume have also been shown in (b).

[14] H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Improved low-count quantitative PET reconstruction with a variational neural network," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3512–22, Nov. 2020.

[15] G. Corda-D'Incan, J. A. Schnabel, and A. J. Reader, "Memory-efficient training for fully unrolled deep learned PET image reconstruction with iteration-dependent targets," *IEEE Trans. Radiation and Plasma Med. Sci.*, 2021.

[16] M. Magnusson, P.-E. Danielsson, and J. Sunnegardh, "Handling of long objects in iterative improvement of nonexact reconstruction in helical cone-beam CT," *IEEE Trans. Med. Imag.*, vol. 25, no. 7, pp. 935–40, Jul. 2006.

[17] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: interpretable, efficient deep learning for signal and image processing," *IEEE Sig. Proc. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[18] H. Nien and J. A. Fessler, "Relaxed linearized algorithm for faster X-ray CT image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2015, pp. 260–3. [Online]. Available: http://web.eecs.umich.edu/~fessler/papers/files/proc/15/web/nien-15-rla.pdf

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014. [Online]. Available: http://arxiv.org/abs/1406.2661

[20] Y. Long, J. A. Fessler, and J. M. Balter, "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1839–50, Nov. 2010.

[21] H. D. Sarkissian, F. Lucka, M. . Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg, "A cone-beam X-ray computed tomography data collection designed for machine learning," *Sci. Data*, vol. 6, p. 215, 2019.

[22] J. A. Fessler, "Michigan image reconstruction toolbox (MIRT) for Matlab," 2016, available from http://web.eecs.umich.edu/ fessler/irt/irt.

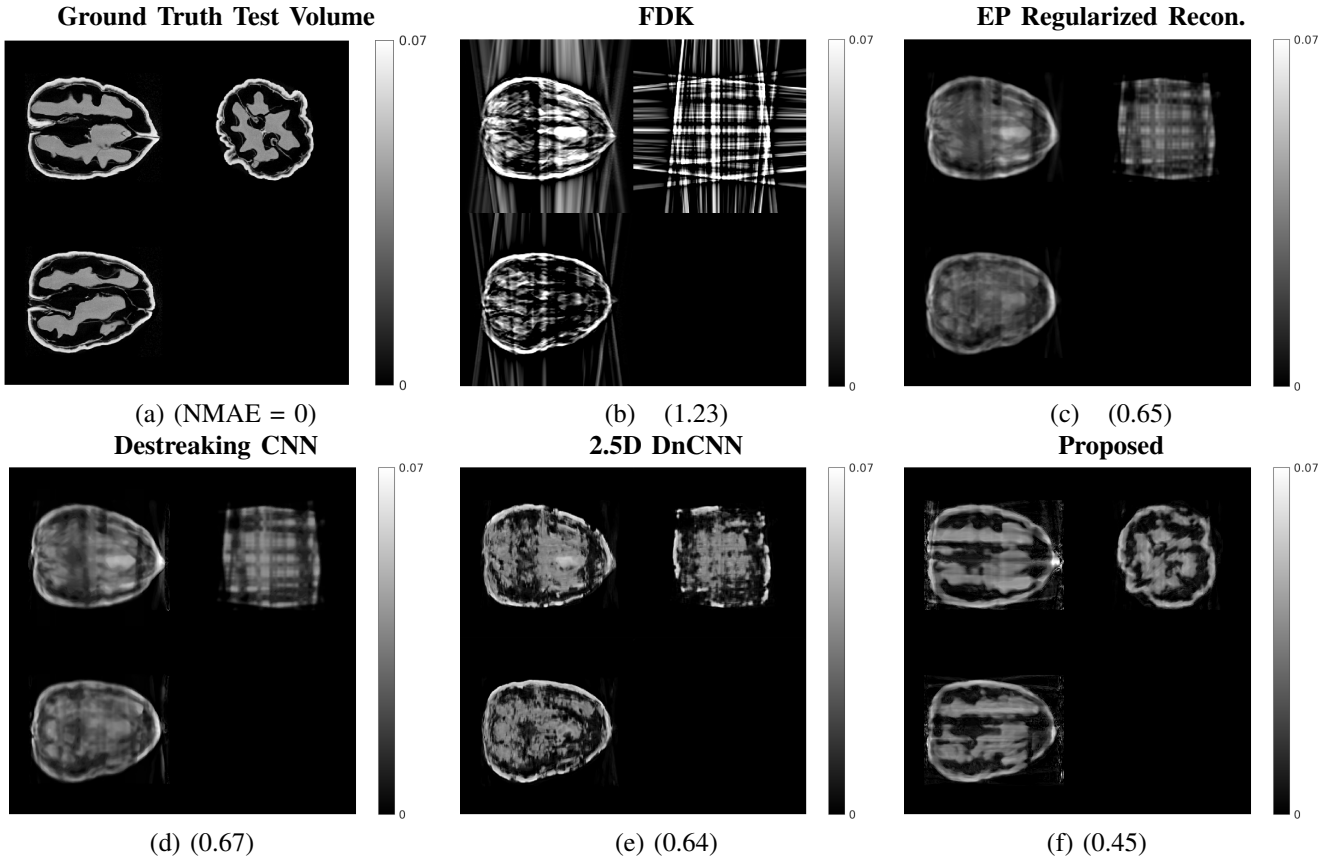


Figure 3. Comparison of the quality of reconstruction of our proposed algorithm (f) for walnut 2 (4 views) in Table II to various reference methods. Each subfigure depicts slices through the center of the walnut volume in the sagittal, coronal and transverse orientations. The normalized mean absolute errors have also been shown underneath each subfigure.

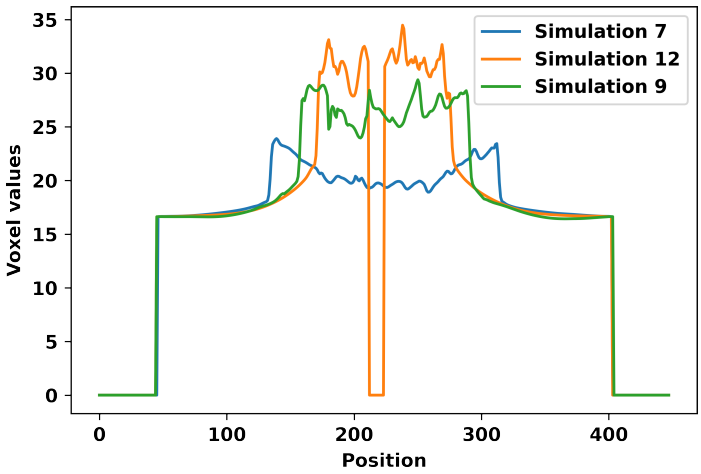


Figure 4. 1-D Profile of central row of central slice of three different Hydrodynamic Simulations.

8 views												
Walnut #	**FDK recon.**		**EP recon.**		**CNN destreaking**		**2.5D DnCNN**		**Proposed 1 stage**		**Proposed 4 stage**	
	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN
1	0.77	0.90	0.45	0.58	0.40	0.59	0.38	0.69	0.33	0.62	**0.26**	**0.54**
2	0.77	0.88	0.45	0.57	0.38	0.58	0.37	0.70	0.32	0.63	**0.25**	**0.53**
3	0.79	0.90	0.49	0.61	0.42	0.62	0.41	0.74	0.38	0.68	**0.30**	**0.58**
4	0.79	0.96	0.45	0.62	0.39	0.62	0.38	0.74	0.34	0.67	**0.27**	**0.58**
5	0.82	0.98	0.48	0.65	0.41	0.66	0.39	0.78	0.35	0.72	**0.27**	**0.61**

Table I

COMPARISON OF THE PERFORMANCE OF OUR PROPOSED METHOD AGAINST FDK RECONSTRUCTION, EDGE-PRESERVING (EP) REGULARIZED RECONSTRUCTION, A SINGLE STAGE OF CNN-BASED DESTREAKING WITHOUT DATA-CONSISTENCY (USING THE SAME ARCHITECTURE AS IN OUR METHOD), A 2.5D DNCNN [13] WITHOUT DATA CONSISTENCY, OUR PROPOSED APPROACH WITH 1 STAGE BUT WITH DESTREAKING CNN TRAINED FOR 160 EPOCHS AND OUR PROPOSED MULTISTAGE RECONSTRUCTION ALGORITHM FOR 8 ACQUIRED PROJECTIONS WITH 4 STAGES WITH DESTREAKING CNN TRAINED FOR 40 EPOCHS EACH. THE METRICS USED FOR PERFORMANCE ARE THE NORMALIZED MEAN ABSOLUTE ERROR (NMAE) AND THE NORMALIZED HIGH-FREQUENCY ERROR NORM (NHFEN).

4 views										
Walnut #	**FDK recon.**		**EP recon.**		**CNN destreaking**		**2.5D DnCNN**		**Proposed**	
	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN
1	1.06	1.08	0.65	0.61	0.65	0.62	0.62	0.71	**0.48**	**0.62**
2	1.14	1.07	0.68	0.60	0.67	0.61	0.64	0.72	**0.51**	**0.62**
3	1.28	1.21	0.66	0.64	0.66	0.65	0.66	0.75	**0.52**	**0.66**
4	1.33	1.32	0.66	0.64	0.65	0.65	0.64	0.78	**0.52**	**0.66**
5	1.31	1.33	0.68	0.69	0.67	0.68	0.65	0.81	**0.53**	**0.71**

Table II

COMPARISON OF THE PERFORMANCE OF OUR PROPOSED METHOD AGAINST FDK RECONSTRUCTION, EDGE-PRESERVING (EP) REGULARIZED RECONSTRUCTION, A SINGLE STAGE OF CNN-BASED DESTREAKING WITHOUT DATA-CONSISTENCY (USING THE SAME ARCHITECTURE AS IN OUR METHOD), A 2.5D DNCNN [13] WITHOUT DATA CONSISTENCY AND OUR PROPOSED MULTISTAGE RECONSTRUCTION ALGORITHM 4 ACQUIRED PROJECTIONS. THE METRICS USED FOR PERFORMANCE ARE THE NORMALIZED MEAN ABSOLUTE ERROR (NMAE) AND THE NORMALIZED HIGH-FREQUENCY ERROR NORM (NHFEN).

[23] Z. Huang, M. Klasky, T. Wilcox, and S. Ravishankar, "Physics-driven learning of Wasserstein GAN for density reconstruction in dynamic tomography," *Appl. Opt.*, vol. 61, no. 10, pp. 2805–2817, Apr 2022. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-61-10-2805

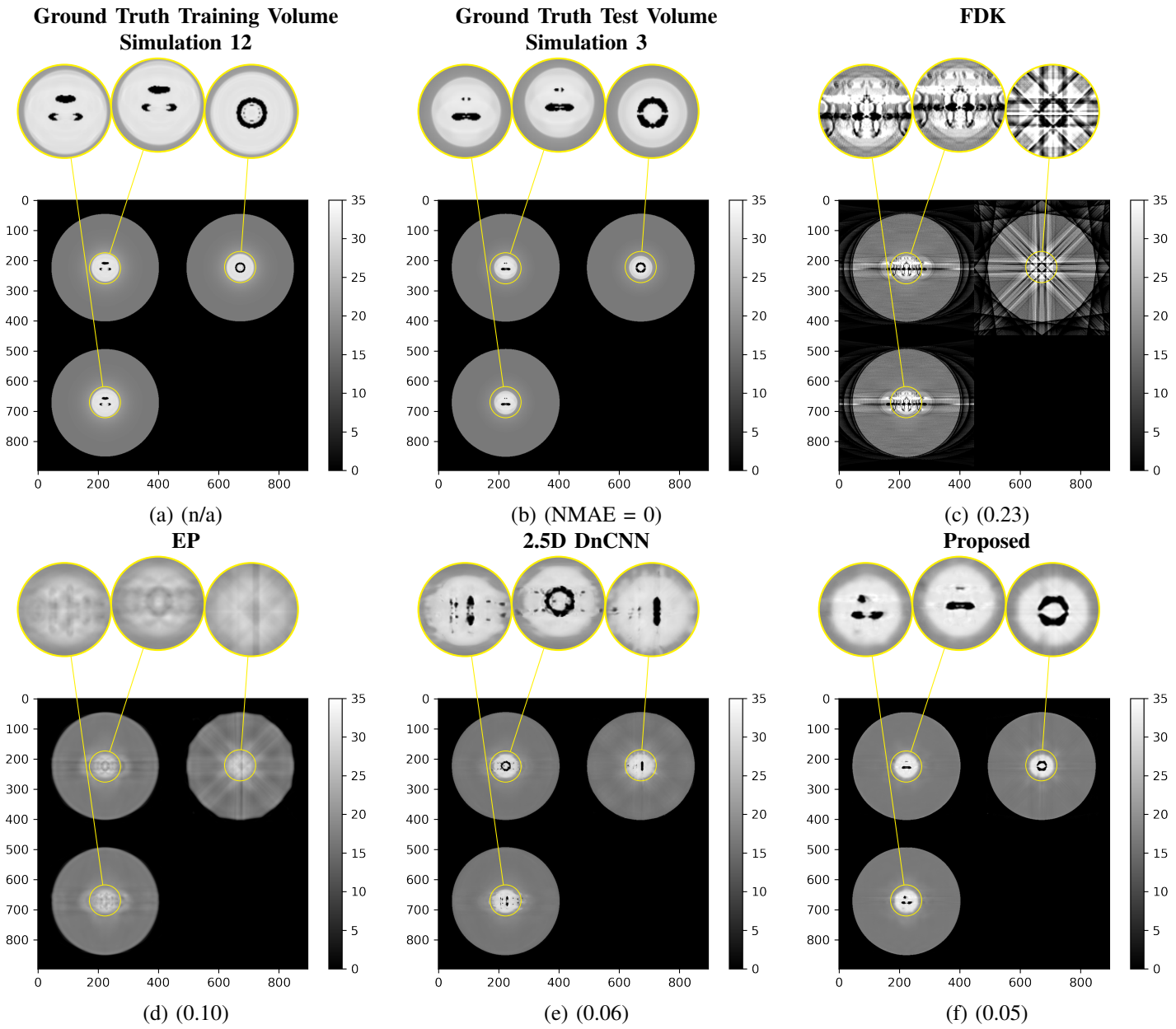
[24] Hertel, E. S. Jr. and Kerley, G. I., "CTH EOS Package: Introductory Tutorial," Sandia National Laboratories, report SAND98-0945, 1998.

[25] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–41, May 2011.

Trained on Simulation 12										
Simulation #	**FDK recon.**		**EP recon.**		**CNN destreaking**		**2.5D DnCNN**		**Proposed**	
	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN
1	0.17	1.07	0.06	0.33	0.11	0.34	0.08	0.34	**0.05**	**0.31**
2	0.21	1.17	0.08	0.37	0.09	0.32	0.07	0.46	**0.05**	**0.3**
3	0.23	1.31	0.1	0.43	0.07	0.39	0.06	0.55	**0.05**	**0.37**
4	0.24	1.43	0.09	0.49	0.08	0.46	0.06	0.6	**0.06**	**0.45**
5	0.22	1.47	0.09	0.54	0.1	0.53	0.11	0.61	**0.09**	**0.52**
6	0.21	1.38	0.09	0.52	0.08	0.52	0.15	0.63	**0.07**	**0.5**
8	0.19	1.31	0.09	0.51	0.08	0.51	0.12	0.55	**0.08**	**0.49**
7	0.2	1.31	0.08	0.47	0.06	0.47	0.14	0.49	**0.05**	**0.45**
9	0.23	1.25	0.09	0.41	0.06	0.33	0.1	0.69	**0.04**	**0.31**
10	0.23	1.54	0.09	0.55	0.07	0.53	0.08	0.68	**0.07**	**0.51**
11	0.24	1.39	0.1	0.48	0.07	0.41	0.05	0.62	**0.05**	**0.4**
Trained on different Simulation 7										
Simulation #	**FDK recon.**		**EP recon.**		**CNN destreaking**		**2.5D DnCNN**		**Proposed**	
	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN	NMAE	NHFEN
1	0.17	1.07	0.06	0.33	0.1	0.47	0.08	0.41	**0.08**	**0.46**
2	0.21	1.17	0.08	0.37	0.09	0.34	0.1	0.46	**0.07**	**0.33**
3	0.23	1.31	0.1	0.43	0.09	0.42	0.12	0.57	**0.08**	**0.4**
4	0.24	1.43	0.09	0.49	0.09	0.46	0.13	0.58	**0.07**	**0.45**
5	0.22	1.47	0.09	0.54	0.07	0.48	0.14	0.57	**0.05**	**0.47**
6	0.21	1.38	0.09	0.52	0.05	0.44	0.13	0.53	**0.04**	**0.43**
8	0.19	1.31	0.09	0.51	0.09	0.46	0.14	0.55	**0.05**	**0.44**
9	0.23	1.25	0.09	0.41	0.1	0.39	0.11	0.57	**0.08**	**0.37**
10	0.23	1.54	0.09	0.55	0.1	0.52	0.14	0.63	**0.07**	**0.5**
11	0.24	1.39	0.1	0.48	0.1	0.45	0.13	0.58	**0.08**	**0.43**
12	0.21	1.27	0.09	0.43	0.1	0.4	0.12	0.54	**0.08**	**0.39**

Table III

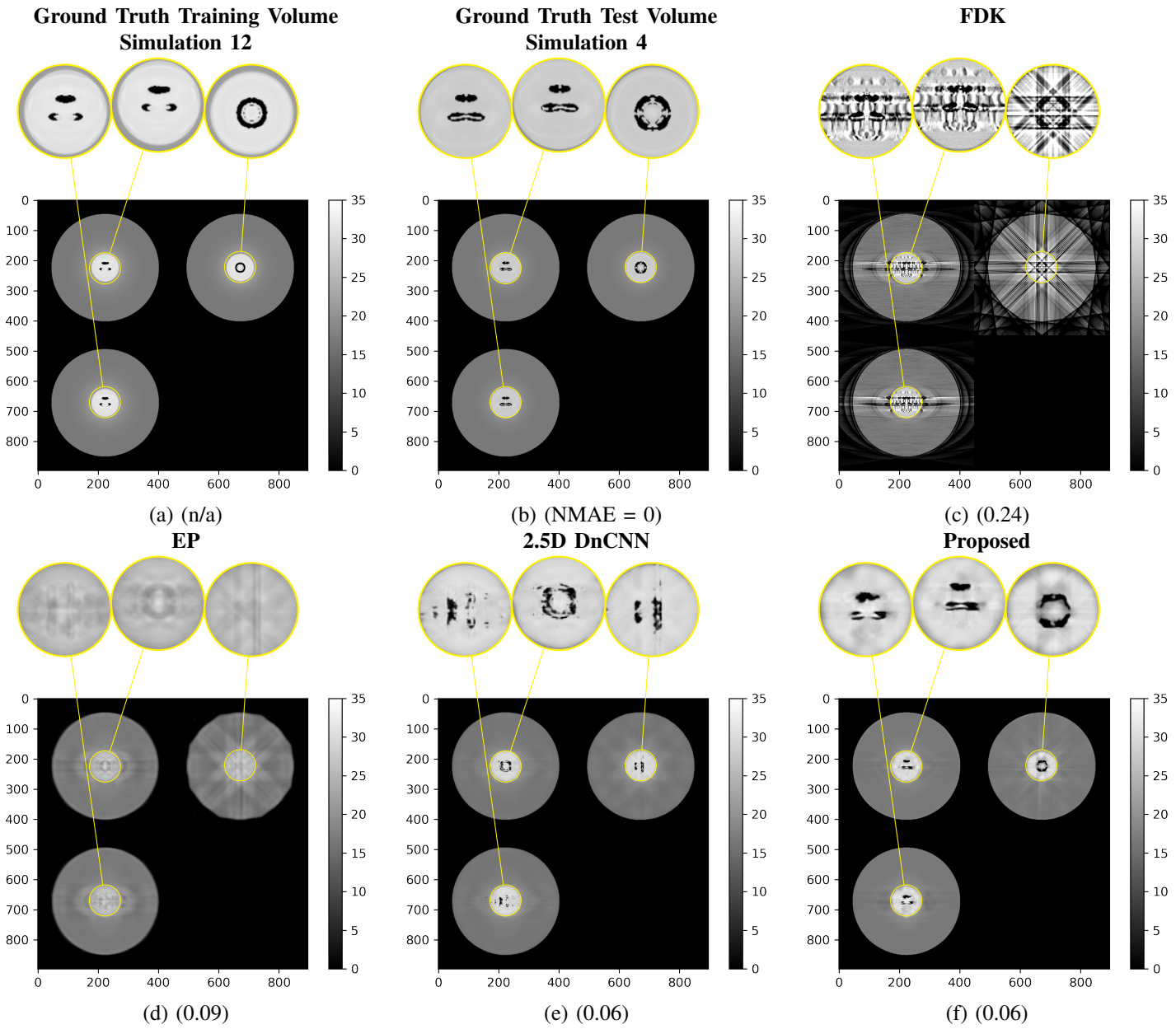
COMPARISON OF THE PERFORMANCE OF OUR PROPOSED METHOD FOR 8 VIEWS AGAINST FDK RECONSTRUCTION, EDGE-PRESERVING (EP) REGULARIZED RECONSTRUCTION AND 2.5D DNCNN [13]. THE TRAINING WAS DONE ON TWO DIFFERENT 3D HYDRODYNAMIC SIMULATIONS. ALL SCORES ARE REPORTED ON A $130 \times 130 \times 130$ PIXEL VOLUME AT THE CENTER OF EACH RECONSTRUCTION.



(a) (n/a) (b) (NMAE = 0) (c) (0.23)

(d) (0.10) (e) (0.06) (f) (0.05)

Figure 5. Comparison of different methods for reconstructing hydrodynamic data from 8 views. All methods were trained on the simulation shown in (a). Zoom-ins of the central portion of each image are provided. The proposed method clearly outperforms the other schemes



(a) (n/a) (b) (NMAE = 0) (c) (0.24)

(d) (0.09) (e) (0.06) (f) (0.06)

Figure 6. Comparison of different methods for reconstructing hydrodynamic data from 8 views. All methods were trained on the simulation shown in (a). Zoom-ins of the central portion of each image are provided. The proposed method clearly outperforms the other schemes